A STRATEGY FOR ESTIMATING TREE CANOPY DENSITY USING LANDSAT 7 ETM+ AND HIGH RESOLUTION IMAGES OVER LARGE AREAS (Part I)

Chengquan Huang, Limin Yang, Bruce Wylie, Collin Homer

USGS EROS Data Center, SAIC, Sioux Falls, SD 57198, USA. Email: huang@usgs.gov

Zhiliang Zhu

USGS EROS Data Center, Sioux Falls, SD 57198, USA.

Abstract

A strategy is developed for estimating tree canopy density at a spatial resolution of 30m. Based on reference data derived from high resolution images, this strategy uses linear regression and regression tree techniques to model tree canopy density from Landsat data. It was tested over three areas of the United States. Regression tree was found more robust than linear regression in all three areas. Mean absolute difference and correlation (*r*) between actual and regression tree predicted canopy density values were about 10% and 0.85 – 0.89, respectively. This strategy will be recommended for use in developing a nation wide tree canopy density data set at a 30m resolution as part of the Multi-Resolution Land Characteristics 2000 project.

Introduction

The Multi-Resolution Land Characteristics (MRLC) consortium was initiated in early 1990s to address the need for consistently developed national and regional land cover data (Loveland and Shaw, 1996). Through this consortium, a 1992-vintage National Land Cover Dataset (NLCD) was developed for the conterminous United States (Vogelmann et al., 2001), and a second generation National Land Cover Database will be developed using 2000-vintage Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images and relevant ancillary data. In addition to a land cover classification, data layers characterizing several key land cover components, including tree canopy density and percent imperviousness, will be developed through the 2000 MRLC effort. These data layers are of increasing relevancy to a variety of scientific and land management applications.

Objectives

Develop a strategy for estimating tree canopy density at intermediate spatial resolutions using ETM+ and high resolution imagery, and assess its applicability to different landscapes in large area applications.

The Strategy

The proposed strategy consists of three key steps: deriving reference data from high resolution images such as IKONOS, DOQQ and QuickBird, developing tree canopy density models, and extrapolating the models spatially using 30m ETM+ images (figure 1). Part of the reference data will be used to evaluate model performance.

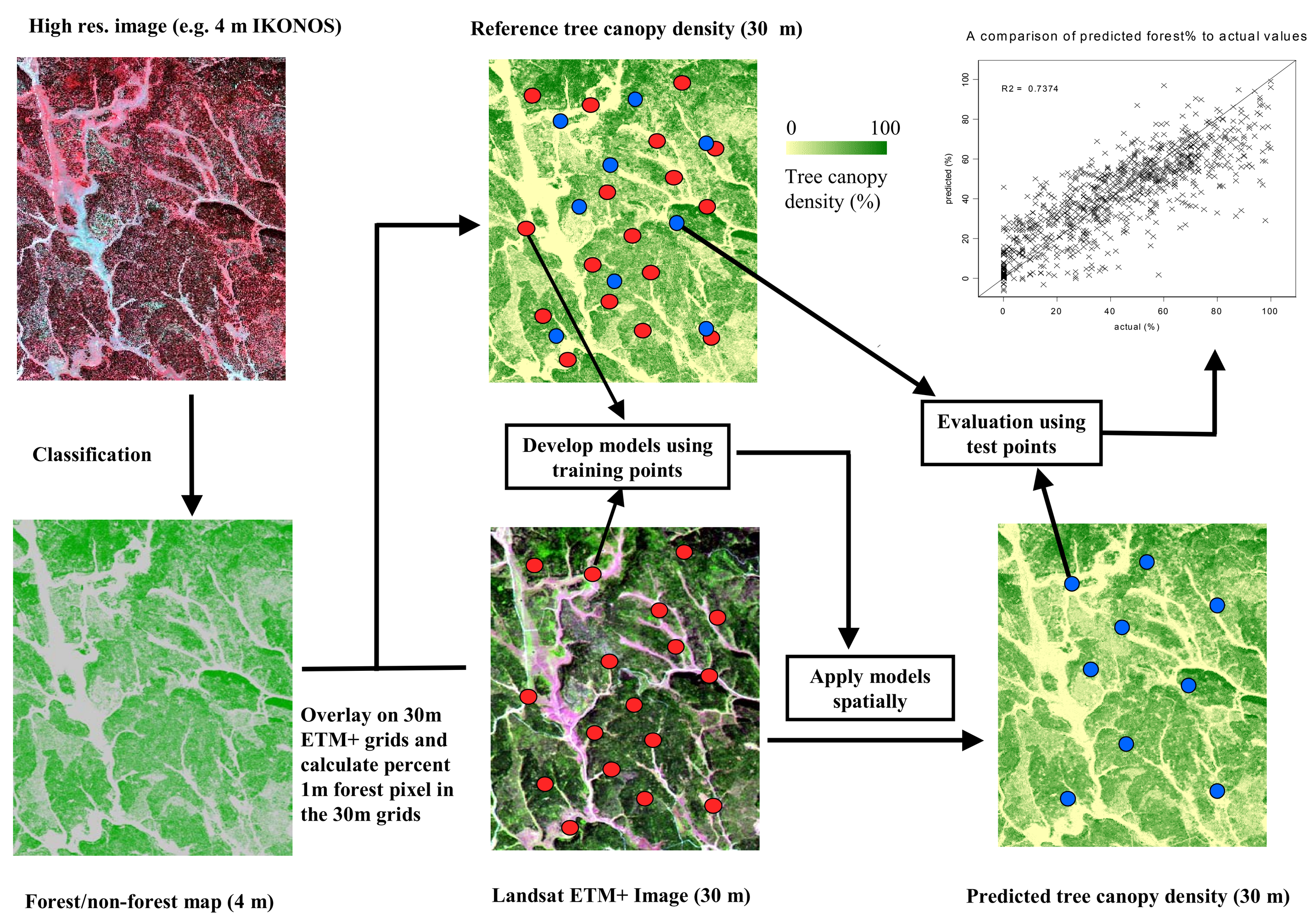

Figure 1. Data flow and procedures for estimating tree canopy density from ETM+ and high resolution images.

ETM+ imagery and preprocessing

Study areas

The proposed strategy were tested in three study areas located in Virginia, Utah/Idaho and Oregon, representing the eastern coast, west semi-arid and pacific northwest landscapes of the United States, respectively (figure 2). Each study area covered two neighboring ETM+ path/rows.

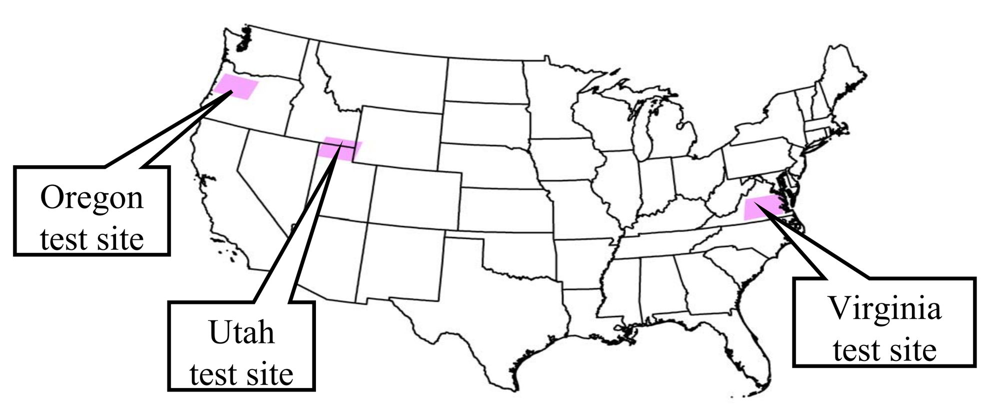

Figure 2. Location of the three study areas

Image preprocessing

The ETM+ images used in this study (table 1) were radiometrically and geometrically corrected using standard methods (Irish, 2000). Location errors due to the impact of terrain relief was corrected using the 1-arc second digital elevation model (DEM) developed by the US Geological Survey (USGS) . The raw images were converted to at-satellite reflectance or brightness temperature according to Irish (2000).

Table 1. Study areas and ETM+ images used in this study

Location	Path	Row	Leaf-on date	Leaf-off date
Virginia	15	34	Jul. 28, 1999	Nov. 17, 1999
	16	34	Jul. 19, 1999	Nov. 8, 1999
Utah	38	31	Aug. 14, 1999	Oct. 17, 1999
	39	31	Jul. 4, 1999	Oct. 24, 1999
Oregon	45	29	Jul. 30, 1999	Dec. 21, 1999
	46	29	Aug. 22, 1999	Dec. 28, 1999

High resolution imagery and reference data development

Reference tree canopy density data at the 30m spatial resolution can be derived from any georeferenced image data with spatial resolutions substantially higher than 30m. In this study, they were derived from 1-m Digital Orthophoto Quadrangles (DOQ) and Space Imaging's IKONOS images. Eight to nine high resolution images were selected for each study area. From each high resolution image a window of 1800 m by 1800 m was identified. These windows were selected to capture spatial, spectral and tree canopy density variations in each study area, and to avoid obvious land cover changes between the high resolution data and the ETM+ images.

Each high resolution image was classified into a forest/non-forest map using a decision tree classifier called C5.0 (Quinlan, 1993). Figure 3 gives the cross-validation estimates of the accuracy of the initial decision tree classifications. Cross-validation is a technique designed to obtain relative objective accuracy estimates without performing a rigorous accuracy assessment (Michie et al., 1994). While these estimates may be inflated for the initial decision tree classifications due to possible spatial auto-correlations between training and test samples, post-classification editing was performed to correct major confusions between forest and non-forest in the initial classifications. Therefore, the accuracy of the final classifications should be close to or better than the cross validation estimates.

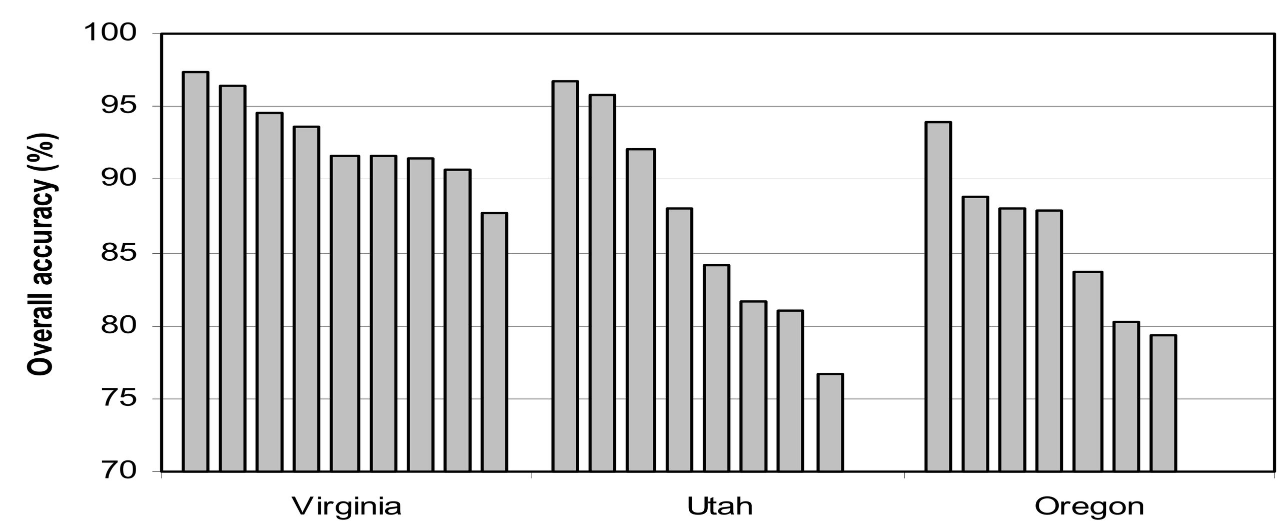

Figure 3. Five-fold cross validation estimates of the accuracy for the decision tree classification of high resolution images. Each bar represents the estimated accuracy of classifying one high resolution image window.

Reference tree canopy density data was derived by calculating the percentage of high resolution forest pixels within the 30m ETM+ grids (figure 4).

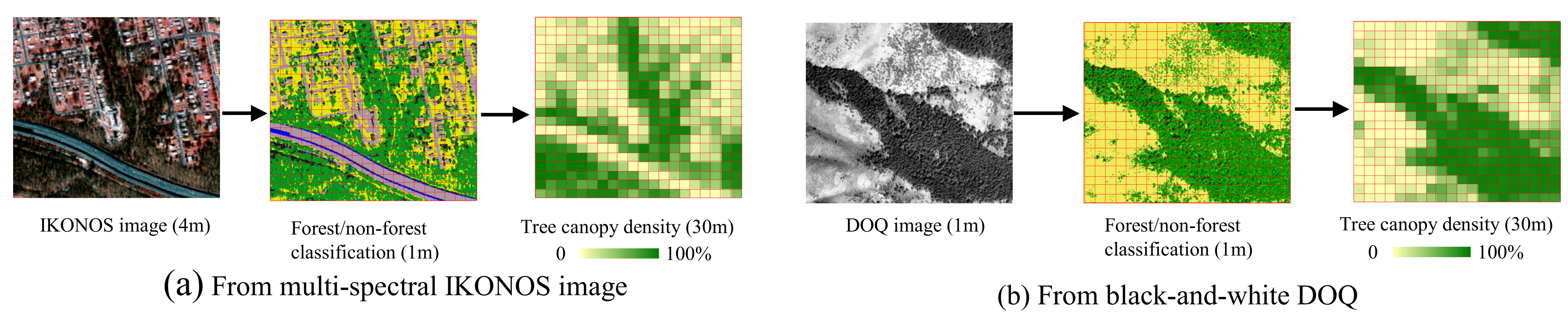

Figure 4. Development of reference tree canopy density data. The red grids are 30m ETM+ grids

A STRATEGY FOR ESTIMATING TREE CANOPY DENSITY USING LANDSAT 7 ETM+ AND HIGH RESOLUTION IMAGES OVER LARGE AREAS (Part II)

Chengquan Huang, Limin Yang, Bruce Wylie, Collin Homer

USGS EROS Data Center, SAIC, Sioux Falls, SD 57198, USA. Email: huang@usgs.gov

Zhiliang Zhu

USGS EROS Data Center, Sioux Falls, SD 57198, USA.

Tree canopy density modeling and evaluation

Training sample selection

The derived 30m reference tree canopy density data were split into training and test data sets as follows (figure 5). Each reference image was divided into 16 equal-sized blocks, twelve of which were randomly selected as training samples and the remaining reserved as test samples. Splitting the reference points by pixel block rather than by pixel can partially reduce spatial auto-correlations between training and test samples, and thus can reduce possible inflation of estimated accuracy (Campbell, 1981). For each study area the training samples from all high resolution image windows were combined to form a training data set and the test samples combined to form a test data set.

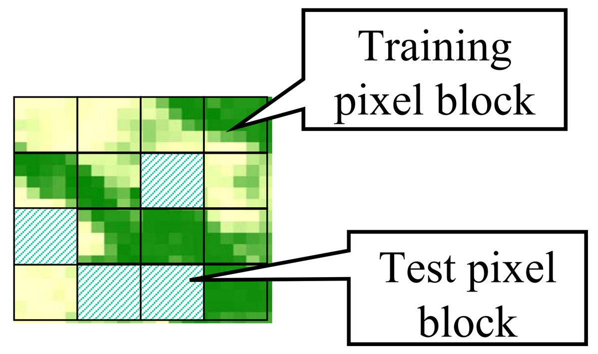

Figure 5. Splitting the reference canopy density data into training and test data sets

Tree canopy density modeling

Tree canopy density was modeled using stepwise linear regression and regression tree techniques. Figure 6 shows an example regression tree. In general, regression tree should be more accurate than linear regression in modeling tree canopy density because it (Huang and Townshend, 2001):

- recursively partitions data samples into subsets
- develops a linear model for each subset
- minimizes the overall residual sum square of error
- can approximate complex nonlinear relationships

The program used in this study, Cubist, is a variant of regression tree.

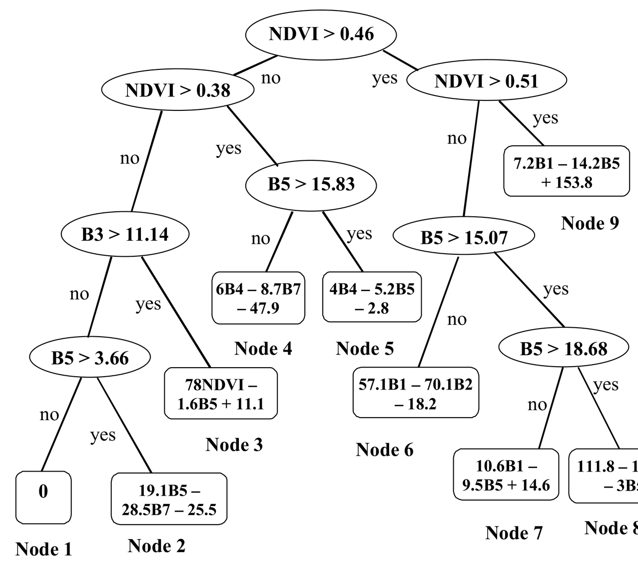

Figure 6. An example regression tree. Tree canopy density is predicted using the linear formula in the terminal nodes.

Model evaluation

The developed models were evaluated using the set aside test data sets. Model performance was measured by the mean absolute difference (*MAD*) and correlation (*r*) between predicted and actual canopy density values (table 2).

Table 2. Mean absolute difference (*MAD*) and correlation (*r*) between predicted and actual canopy density values on independent test samples. The unit of *MAD* is tree canopy density in percentage.

Study area	Regression tree model		Linear regression model	
	MAD (%)	*r*	*MAD* (%)	*r*
Virginia	11.65	0.89	13.15	0.83
Utah	9.92	0.85	10.14	0.70
Oregon	10.98	0.87	11.93	0.80

The regression tree had substantially lower *MAD* and higher *r* values than linear regression in all three study areas. Its performances were relatively consistent over the three different areas, indicating the general applicability of the proposed strategy to estimating tree canopy density over large areas. The residual errors are mostly likely due to the complex and highly variable nature of mixings between tree canopy and non-canopy surface materials. Other sources include noises in reference data and the ETM+ images. The former may arise from errors in classifying the high resolution images, partial canopy cover pixels in the high resolution images, and temporal discrepancies and residual registration errors between high resolution images and ETM+ data. Modeling error will likely decrease if such uncertainties can be reduced.

For each study area, tree canopy density was estimated for the entire study area using the developed regression tree models (figure 7).

Conclusions

- A strategy was developed for deriving tree canopy density at a spatial resolution of 30m using ETM+ and high resolution images. This strategy can be used with IKONOS, QuickBird and DOQQ data that have pixel sizes much smaller than 30 m.
- The regression tree program used in this study consistently outperformed stepwise linear regression in modeling tree canopy density from Landsat 7 ETM+ images.
- The mean absolute differences and correlation (*r*) between actual and regression tree predicted canopy density values were around 10% and 0.85 – 0.89, and were relatively consistent over the three study areas.
- The errors likely will be reduced by assigning 0% tree canopy cover to areas such as large agricultural fields, which very unlikely will have any tree cover. This may be achieved by using a conservative non-forest mask.
- With the increasing availability and decreasing cost of both high resolution and ETM+ images, this strategy likely will be applicable to many regions of the world.

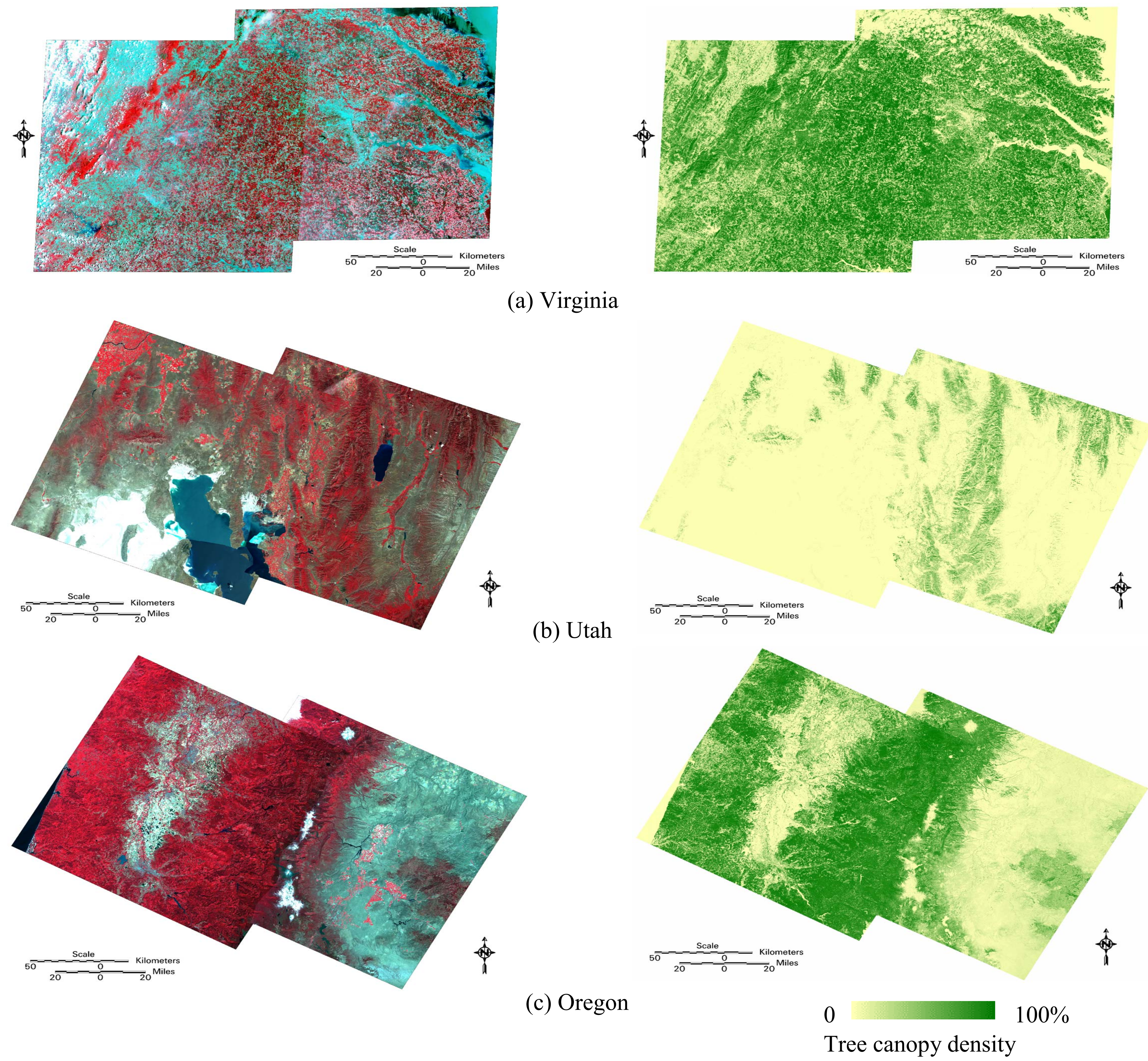

Figure 7. Standard false color leaf-on ETM+ images (left) and tree canopy density images (right) predicted using the regression tree models.

References

J. B. Campbell, "Spatial correlation effects upon accuracy of supervised classification of land cover," *Photogrammetric Engineering & Remote Sensing*, Vol. 47, No. 3, pp. 355-363, 1981.

M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sensing of Environment*, Vol. 61, No. 2, pp. 399-409, 1997.

C. Huang and J. R. G. Townshend, "A stepwise regression tree for nonlinear approximation: applications to estimating subpixel land cover," *International Journal of Remote Sensing*, accepted, 2001.

R. R. Irish, *Landsat 7 science data user's handbook*, Report 430-15-01-003-0, National Aeronautics and Space Administration, http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html, 2000.

T. R. Loveland and D. M. Shaw, "Multiresolution land characterization: building collaborative partnerships." In *Proceedings of the ASPRS/GAP Symposium -- Gap analysis: a landscape approach to biodiversity planning,* Charlotte, North Carolina, pp 83-89, 1996.

D. Michie, D. J. Spiegelhalter, and C. C. Taylor (Editors), 1994, *Machine learning, neural and statistical classification* (New York: Ellis Horwood), 289 pp.

J. R. Quinlan, *C4.5 programs for machine learning,* Morgan Kaufmann Publishers, San Mateo, California, p. 302, 1993.

J. E. Vogelmann, S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. V. Driel, "Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources," *Photogrammetric Engineering & Remote Sensing*, Vol. 67, No. 6, pp. 650-662, 2001.